Feeding the world one open access plant phenotype image at a time

Dr. Sotirios A. Tsaftaris
(Sotos)

http://tsaftaris.com

THE UNIVERSITY of EDINBURGH

The Alan Turing Institute

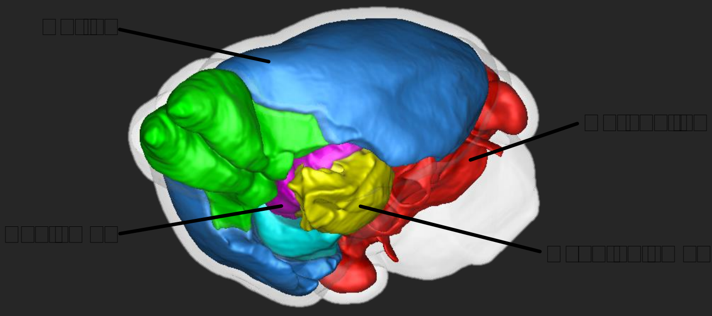
Institute of Digital Communications, School of Electrical Engineering

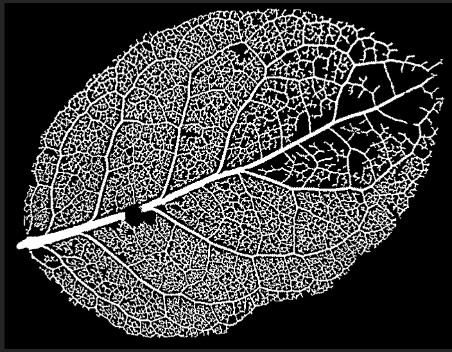
An outline

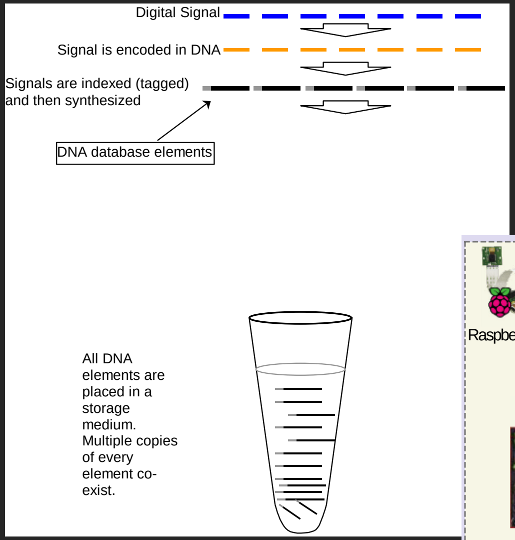
- About us
- Phenotiki and affordable phenotyping
  - Powered by affordable open hardware
  - Smart, machine-learning, open software
- Open data
- Lessons (software lakes for data swans)

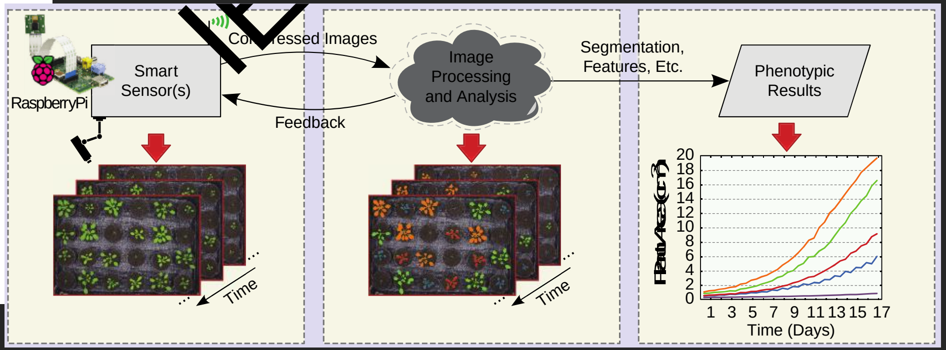
The Lab

10 random images from papers









(Phenotype)

- Appearance / behavioral variability in organisms

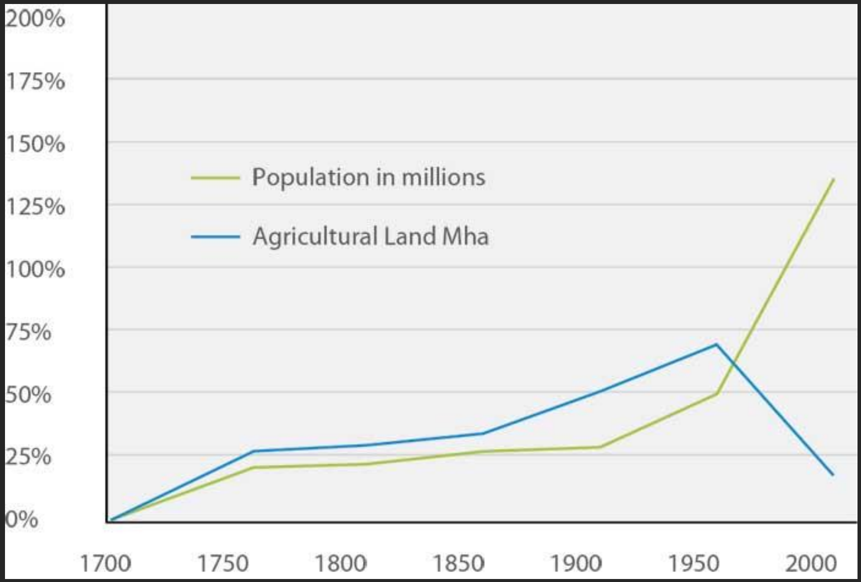
  e.g., how we look,
  how we respond to stress

- Genotype x Environment x Random Variations

Phenotype

Phenotyping is important

- Population increases, resources decrease, climate change
- We need sustainable agriculture



- Phenotyping: measuring traits & reactions
- Missing link to other omics technologies

Phenotyping is (was) a true Bottleneck

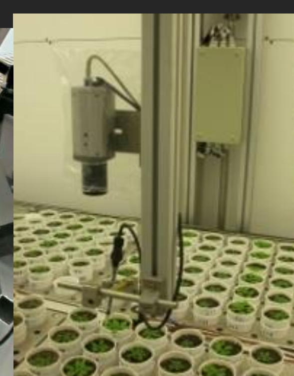
- Growth rate
- Flowering time
- Seed set
- Seed shape
- Leaf shape
- Colour changes
- Root density
- Nutrient utilisation
- Light sensitivity
- ...

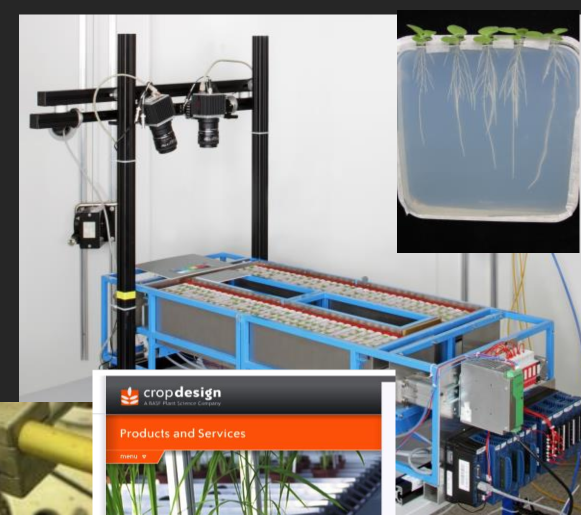
Collecting phenotypes manually is hard!

Use cameras and images to help collection

High throughput phenotyping

- Automated imaging and semi-automated analysis
  - Automation to collect imaging data
  - However, customized and costly solutions





Affordable Plant Phenotyping with Phenotiki

- Really affordable sensor(s) <200€
- Distributed sensing and analysis
- Robust analysis software running on a cloud infrastructure
  + Easy maintenance / deployment, no software needed
  + Transparent to the user
  + Expandable to other organs/plar


the plant journal SEB

Technical Advance

Phenotiki: An open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants

Massimo Minervini, Mario Valerio Giuffrida, Pierdomenico Perata, Sotirios A. Tsaftaris

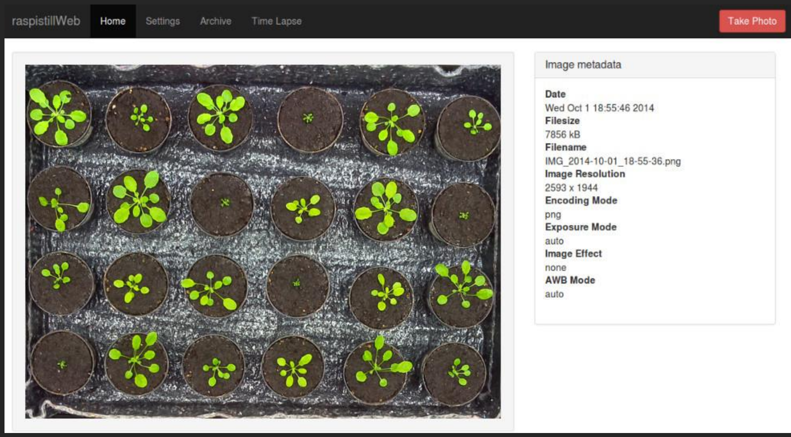
Accepted manuscript online: 9 January 2017

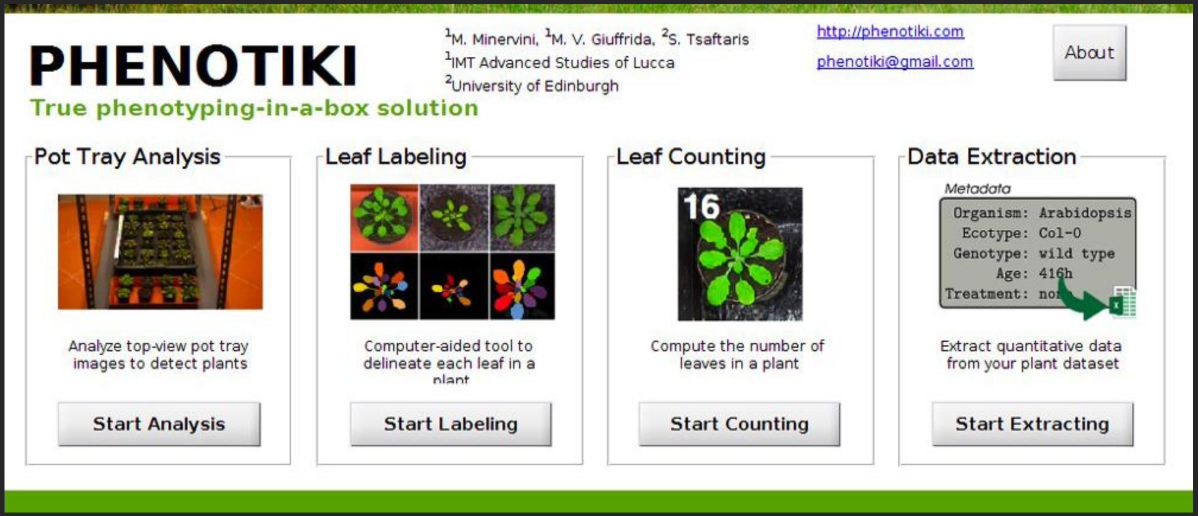
http://phenotiki.com

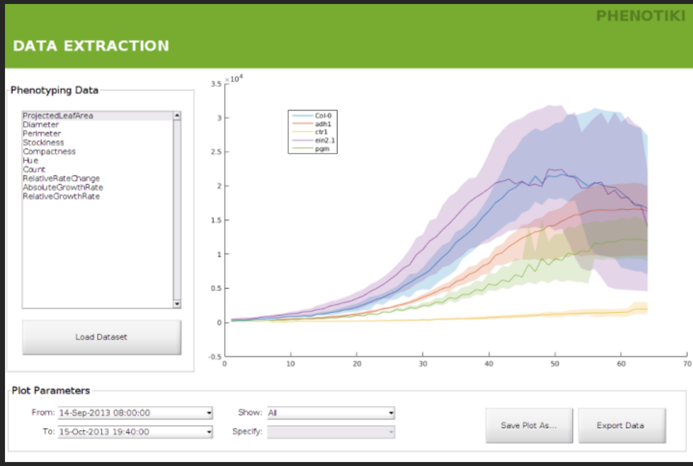
iPlant Collaborative™
NSF
SEVENTH FRAMEWORK PROGRAMME
BBSRC bioscience for the future

3 steps

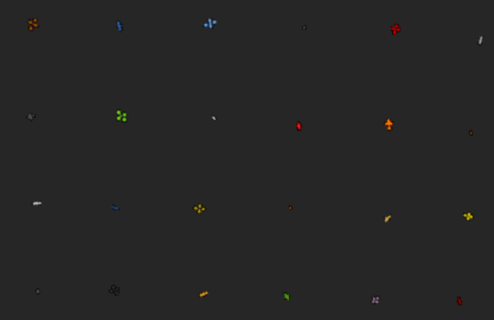
- Setup the sensor <200£
- Connect it to the internet
- Analyze the 2D data
  - On a workstation
  - On the cloud [iPlant]



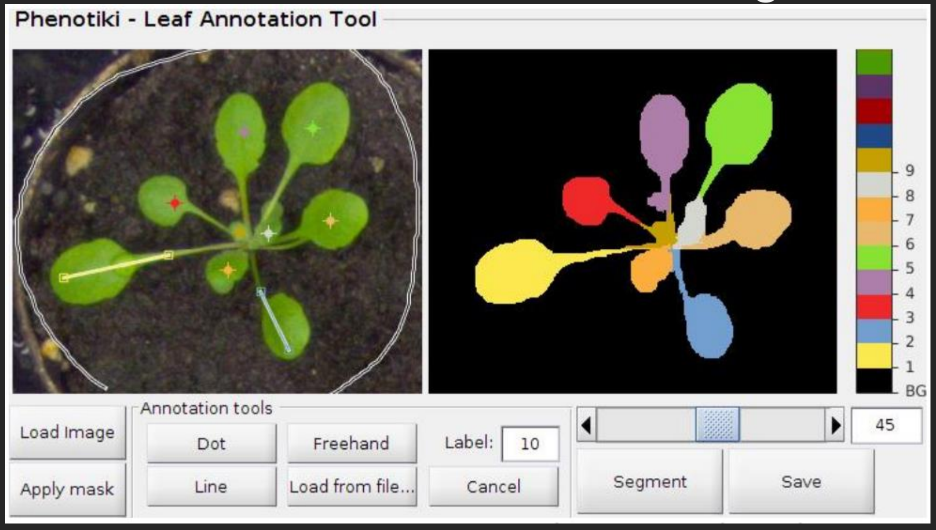




The secret: Machine Learning Algorithms



- Algorithms rooted in machine learning
  - Robust to changing environment (different labs)
  - Learn from user interaction
- Once we teach the algorithms
  - Fully automated plant growth
  - Fully automated leaf counting (1st ever in 2D)
  - Semi-automated leaf segmentation



Traits:
- Projected Leaf Area (PLA)
- Diameter
- Perimeter
- Compactness
- Stockiness
- Leaf count
- Relative Growth Rate
- Color

Getting the phenotypes: the **true** bottleneck

- Sometimes easy… (rosette area)
- Most times hard… challenging:
  - content (moss, drought, water)
  - phenotype (leaf, flower)

- Particularly when we have to image different things in different settings

IEEE SIGNAL PROCESSING MAGAZINE [126] JULY 2015 1053-5888/15©2015IEEE

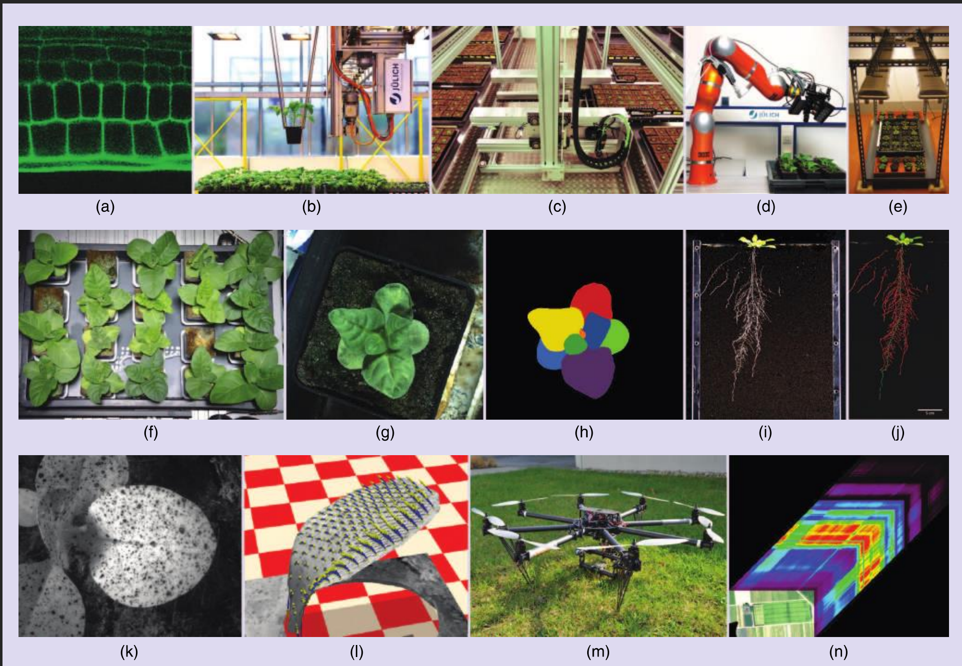
[applications CORNER]

Massimo Minervini, Hanno Scharr,
and Sotirios A. Tsaftaris

Image Analysis: The New Bottleneck in Plant Phenotyping

While the bottleneck was previously the equipment (the hardware), it is now the analysis (the software). There is a

ments. Experts (from biology as well as data analysis) now agree that the analysis of imaging data is currently the weakest, or even the missing, link due to the



A bottleneck that analysis together with machine learning (ML) can help address

Trends in Plant Science

CellPress

Letter

Machine Learning for Plant Phenotyping Needs Image Processing

Sotirios A. Tsaftaris,[1,*] Massimo Minervini,[2] and Hanno Scharr[3]

are related to how we perceive and analyze an object of interest, such as segmentation, detection, tracking, and many others).

When this is not the case, plant segmentation can be extremely complex because here the objects of interest may touch and overlap each other (known as occlusion), as in Figure 1B. In the open field [6] this becomes exceedingly more complex: light variations, plant movements due to wind,

For example, in drought-tolerance studies one can rely on the overall amount of green or yellow pixels as potential features. However, this simple approach may not always allow us to discriminate between stressed and not stressed plants. It is well known in machine learning that finding good features for the application at hand is intrinsic to an effective use of learning approaches (even sophisticated ones). Thus, image processing is key to obtaining accurate and reliable phe-

- ML: teach machines from diverse examples
  - Give images & desired output (trait) ➔ let algorithms decide
  - E.g. Contrast this with deciding (by eye) thresholds to delineate plants for background plus cleaning for PLA

However developing ML algorithms needs data

- When we started in 2011 there was no open data available
  - Despite major academic players (and companies) having made significant contributions in the area
  - Our plant scientists collaborators did not have imaging equipment in place yet
- Luckily we were developing Phenotiki
  - We were doing our own experiments
  - We were collecting our own data
  - We were free to do whatever we wanted with the data

About 2 years ago

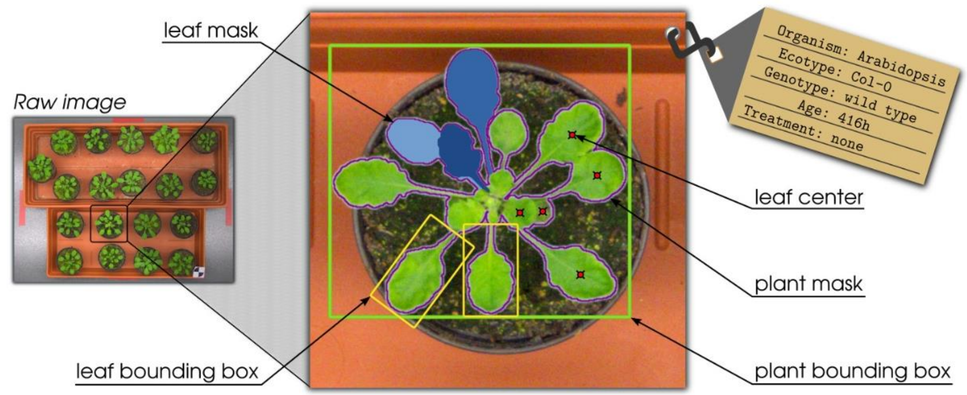
In this paper we present a collection of benchmark datasets for the development and evaluation of computer vision and machine learning algorithms in the context of plant phenotyping. We provide annotated imaging data and suggest suitable evaluation criteria for plant/leaf segmentation, detection, tracking as well as classification and regression problems. The Figure symbolically depicts the data available together with ground truth segmentations and further annotations and metadata.

Pattern Recognition Letters xxx (2015) xxx–xxx

Finely-grained annotated datasets for image-based plant phenotyping

Massimo Minervini [a,*], Andreas Fischbach [b], Hanno Scharr [b], Sotirios A. Tsaftaris [a,c]

[a] *Pattern Recognition and Image Analysis Research Unit, IMT Institute for Advanced Studies, 55100 Lucca, Italy*
[b] *Institute of Bio- and Geosciences: Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany*
[c] *Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh EH9 3JL, UK*



- Purposely in a vision journal
- Data + evaluation routines
- If adopted we can track progress

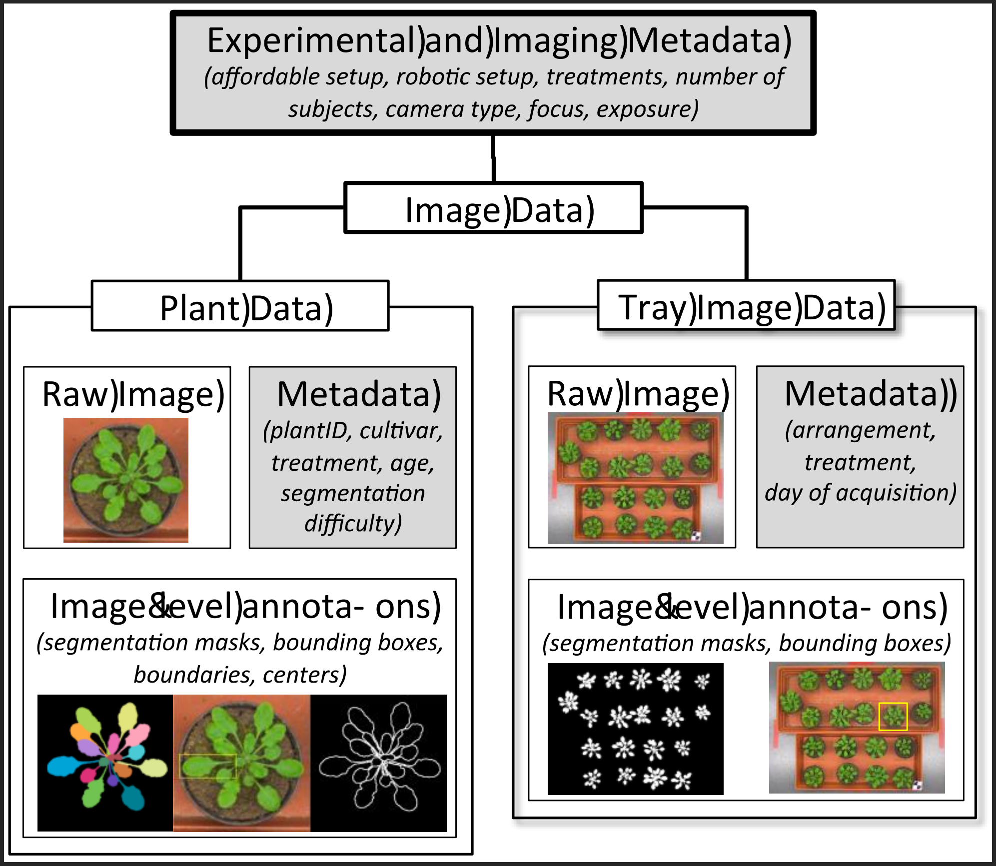
http://www.plant-phenotyping.org/datasets

Collected data

- Collected
  - different experiments of Arabidopsis
  - Tobacco
  - Different cameras
  - Different setups
  - Different illumination
- Recruited annotators and designed an annotation whitepaper

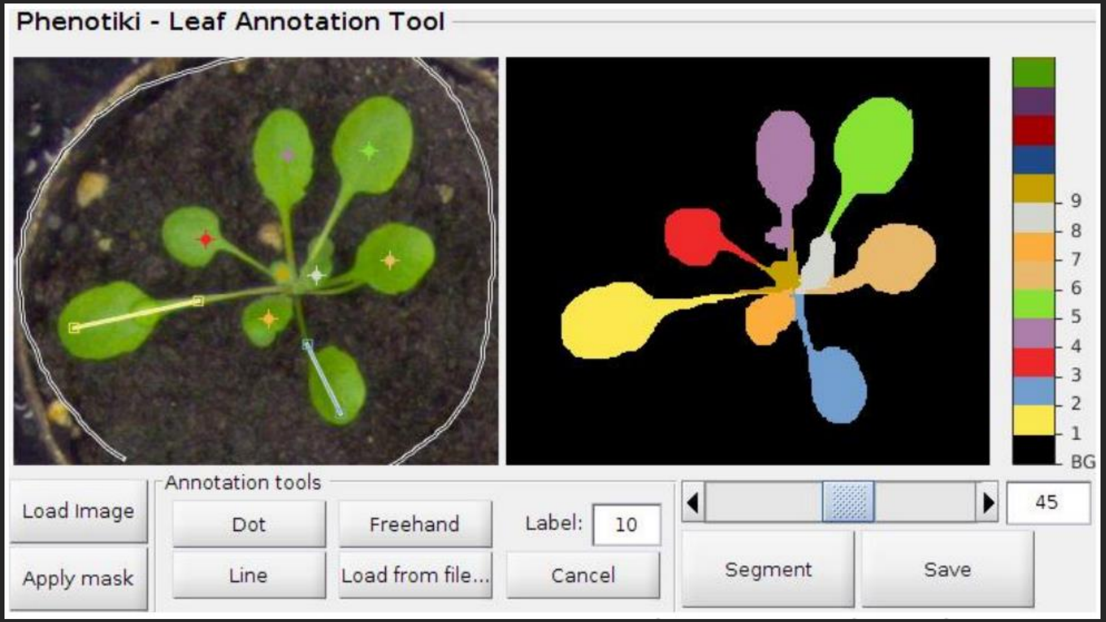
An annotation hierarchy

- We setup a hierarchy
- Minimize annotation effort
- Metadata: easy
- Image-level annotations: harder
- Found the lowest element to annotate (the leaf)➔ derived different annotations from that



Build a tool to delineate leaves to minimize variation and time

- Remarkable segmentation results (≈ 97% **accuracy**)
- **Easier** and **faster** (1 min) vs. raster graphics editors (30 min)



- Publicly **available** software tool and source code
  - Web page: http://www.phenotiki.com
  - GitHub repository: https://github.com/phenotiki/LeafAnnotationTool

Minervini, et al BMVC 2015.

HOW DATA HAVE BEEN USED

Collating expertise

Machine Vision and Applications
DOI 10.1007/s00138-015-0737-3

CrossMark

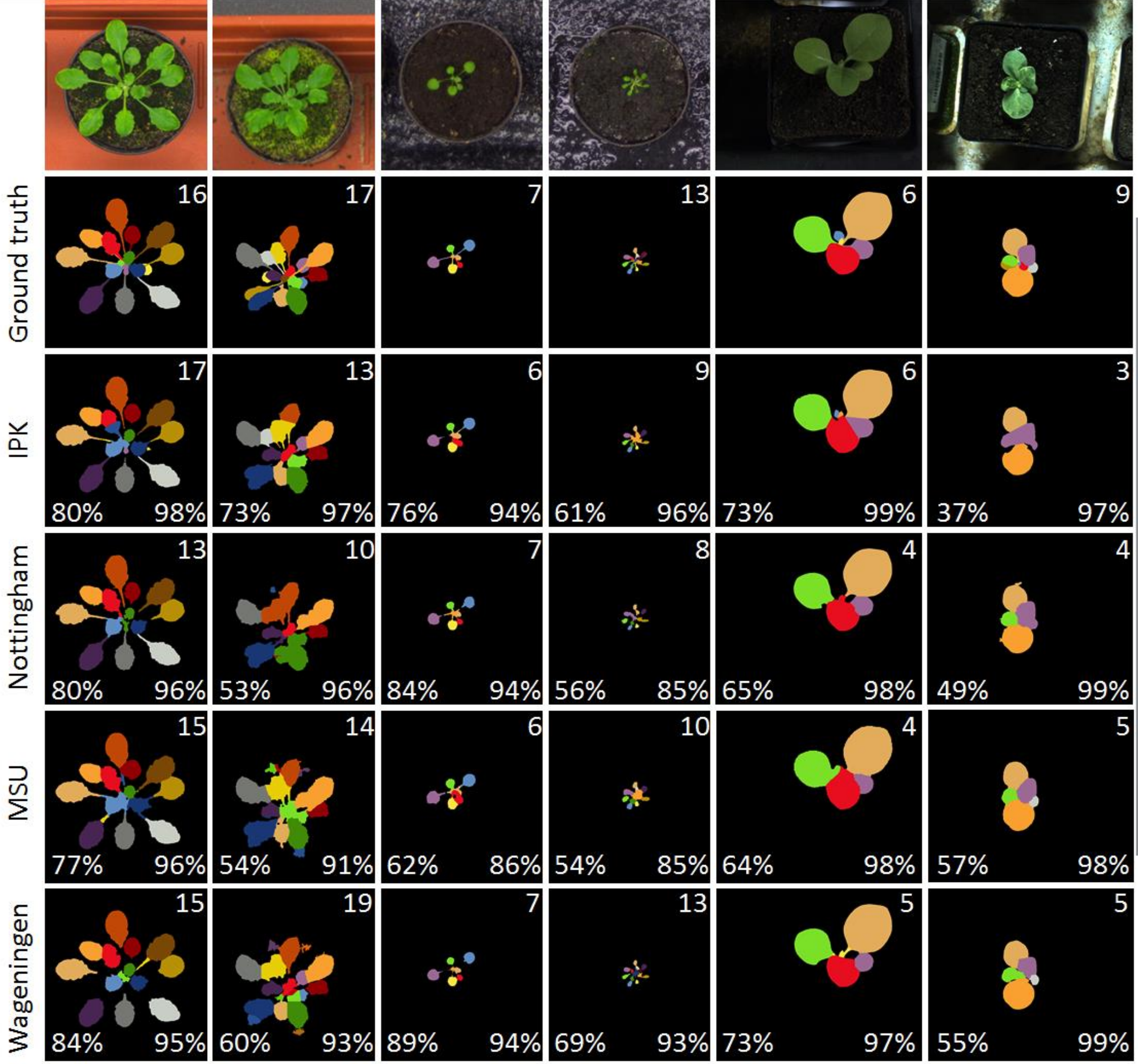
SPECIAL ISSUE PAPER

Leaf segmentation in plant phenotyping: a collation study

Hanno Scharr[1] · Massimo Minervini[2] · Andrew P. French[3] · Christian Klukas[4] · David M. Kramer[5] · Xiaoming Liu[6] · Imanol Luengo Muntión[3] · Jean-Michel Pape[4] · Gerrit Polder[7] · Danijela Vukadinovic[7] · Xi Yin[6] · Sotirios A. Tsaftaris[8,9]

- Benefits of having open data
- Organized challenges (2014,2015,2017)
- 4 groups around the world; compared on the same dataset

Brief Results



Rosette segmentation ok

Leaf

Depends:

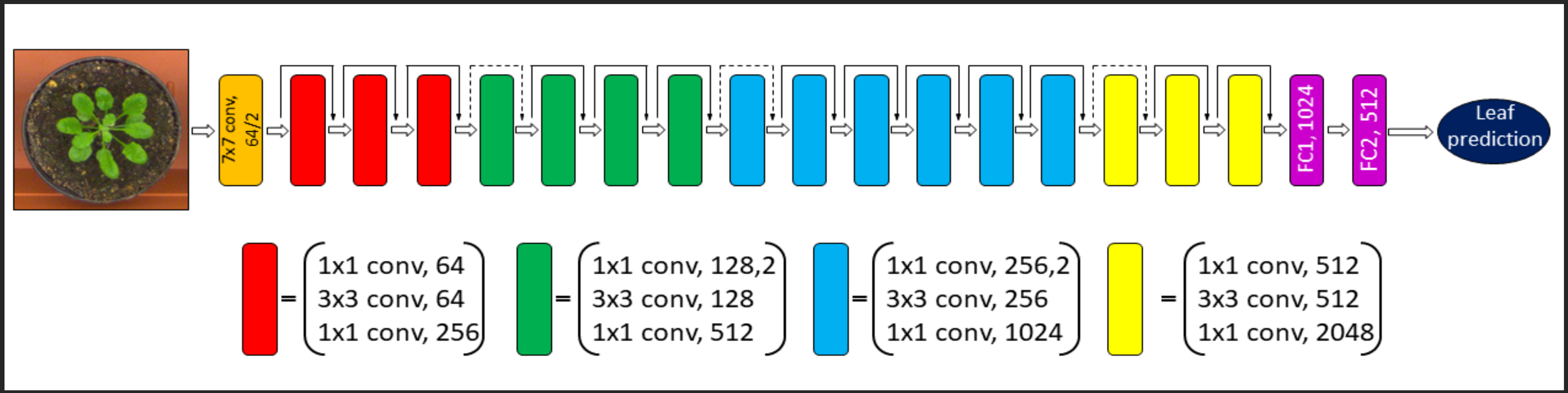
- Young vs mature
- Occlusion a main problem

Thankfully science evolves

	SBD ″	|DiC| #
RIS+ CRF [19]	66.6 (8.7)	1.1 (0.9)
MSU [20]	66.7 (7.6)	2.3 (1.6)
Nottingham [20]	68.3 (6.3)	3.8 (2.0)
Wageningen [26]	71.1 (6.2)	2.2 (1.6)
IPK [14]	74.4 (4.3)	2.6 (1.8)
PRIAn [6]	-	1.3(1.2)
Ours	84.9 (4.8)	0.8 (1.0)

Ren & Zemel End-to-End Instance Segmentation and Counting with Recurrent Attention CVPR 2017

Built state of the art algorithms

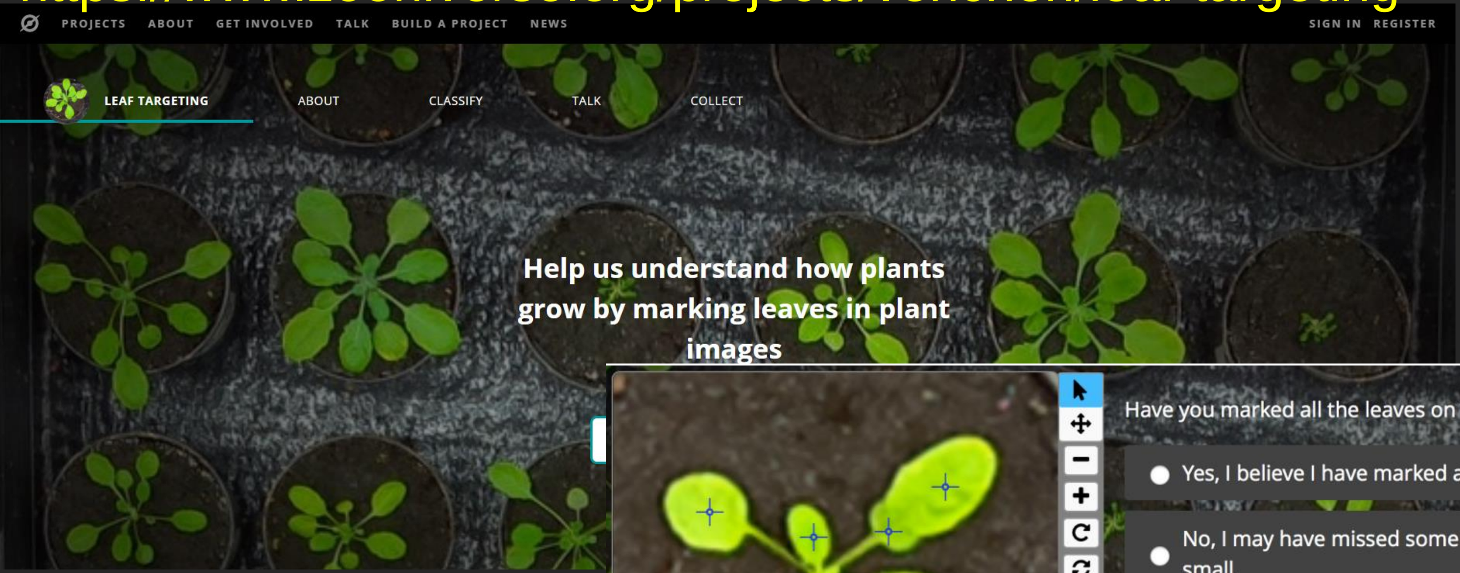


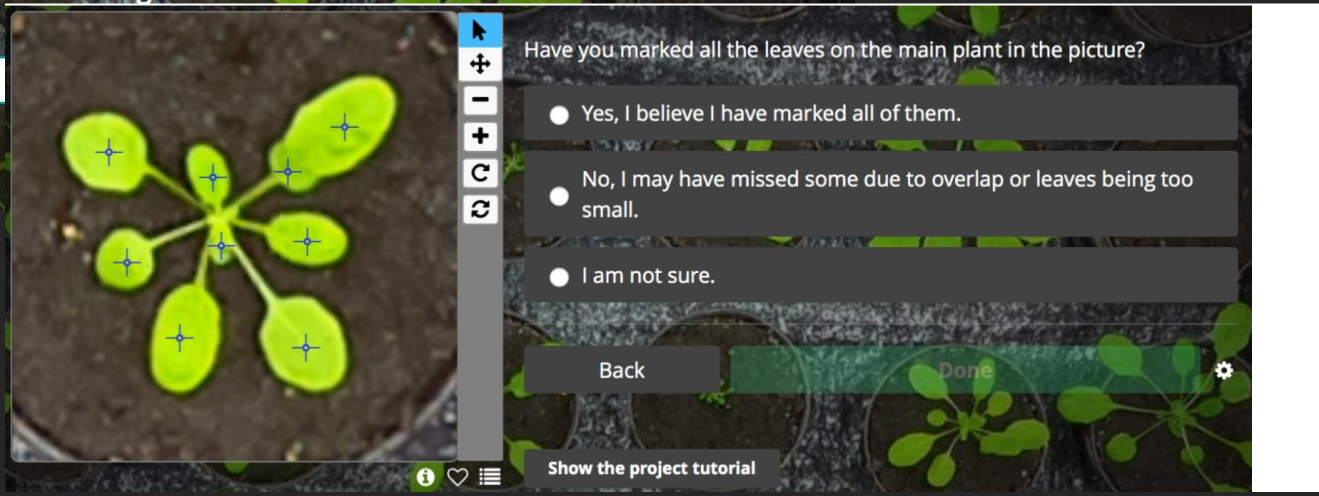
- Deep learning approach direct image to count (for any plant)
- Winner of the 2017 Leaf Counting Challenge (CVPPP 2017)
- Benefits by pooling data sources together
  - Extension to multimodal data [e.g. fluorescence, depth, infrared] forthcoming
- Results improve with more sources and more labeled data
- Results improve with synthetic data

Getting more labelled data

- 20000 annotated plants in 3 months

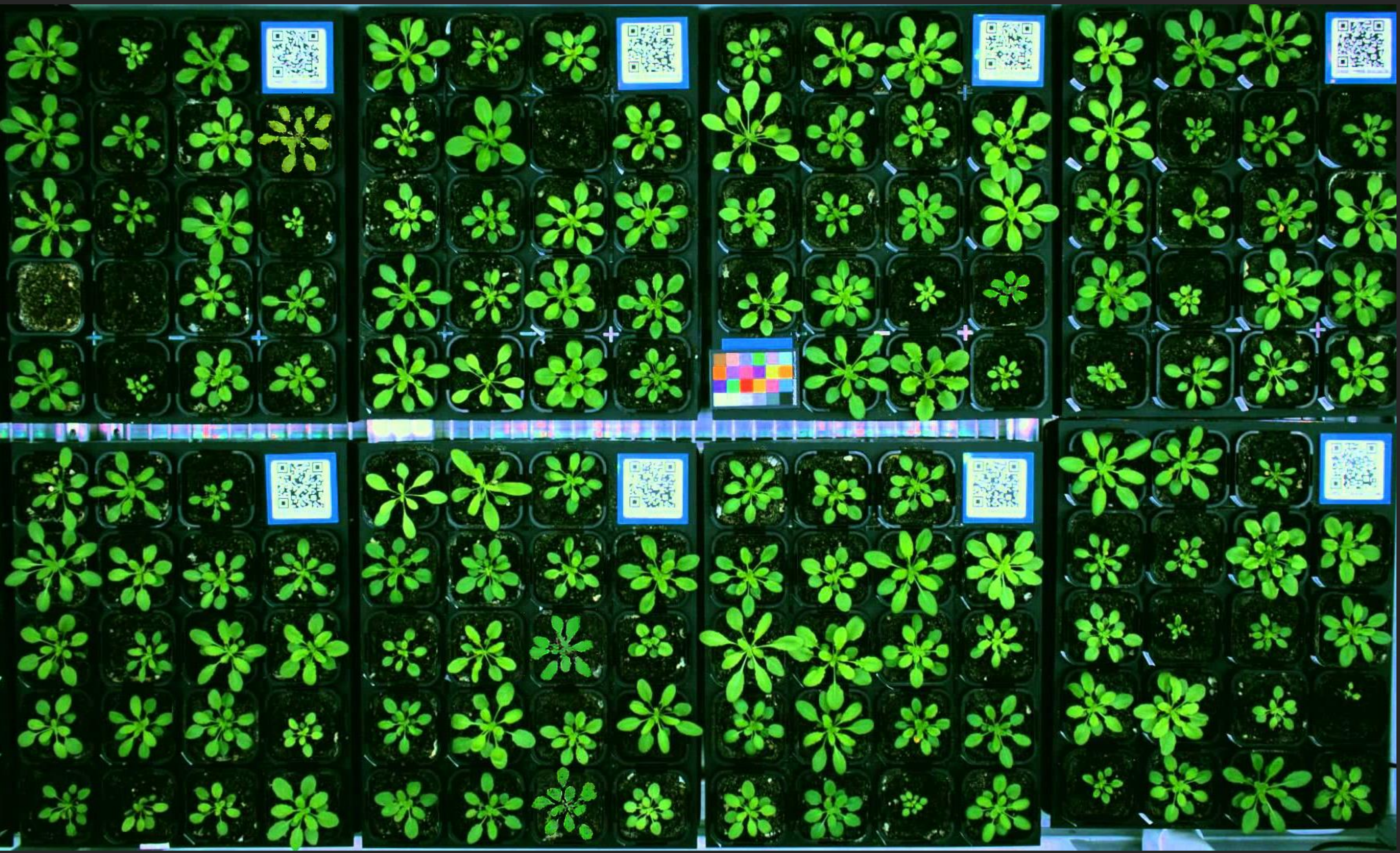
https://www.zooniverse.org/projects/venchen/leaf-targeting





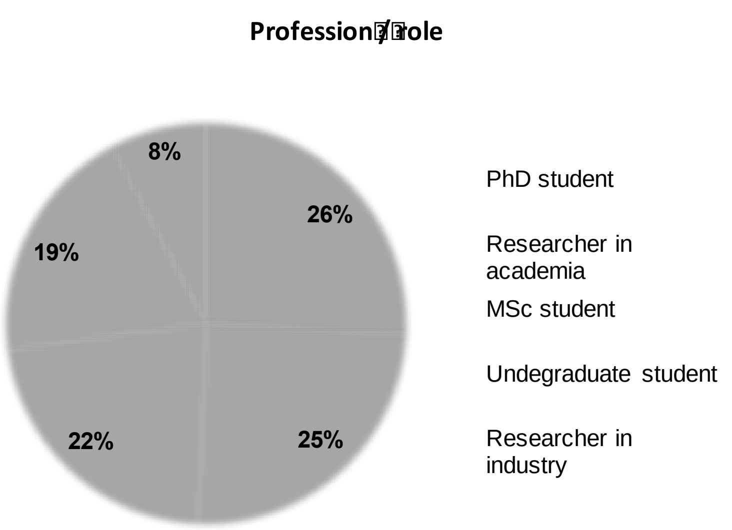
Giuffrida et al under review Plant Methods.

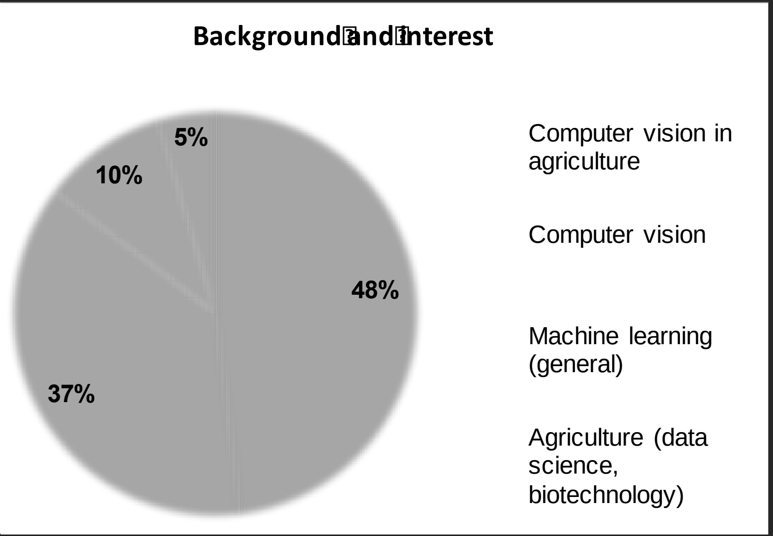
Creating synthetic data: Can you tell the fake from the real?



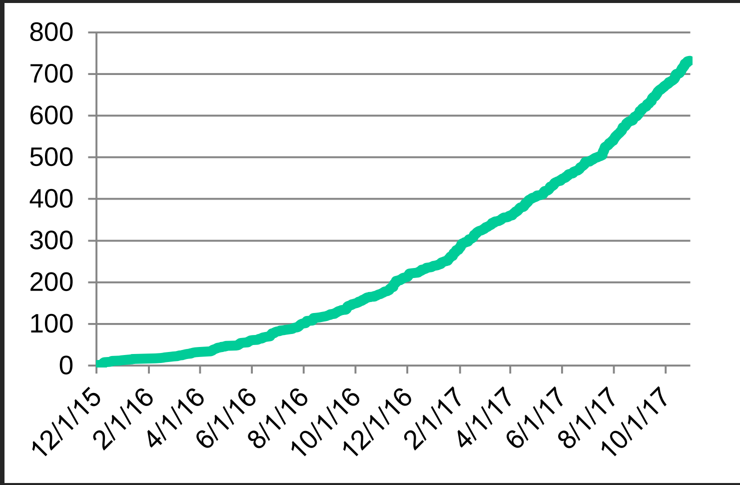
Giuffrida et al CVPPP @ ICCV 2017

How others use the data (2 years after)





- 23 citations
- ~800 downloads
- We succeeded in attracting new CV/ML scientists
  - Most are researchers and students
  - most students are in computer science
- 7 papers in deep learning using the data […]



Lessons and what we need in the future

- Going from benchmark small scale open data to…

- "Data lakes for software swans"
  - Ways to curate of available data/feedback
  - Communicate with users
  - A common framework to collect data for the purpose of developing and testing algorithms from a variety of sites, systems etc
  - Ways to collect annotations for tasks
  - Ways to obfuscate biological knowledge
    - Focus on underlying vision problem [openly]
    - Test and evaluate algorithms then merge with biological knowledge "privately"

Thank you

Funding:

NIH 4R01HL091989
NIH 1R01HL136578

EP/N510129/1

EP/P022928/1

BB/N02334X/1
BB/P023487/1

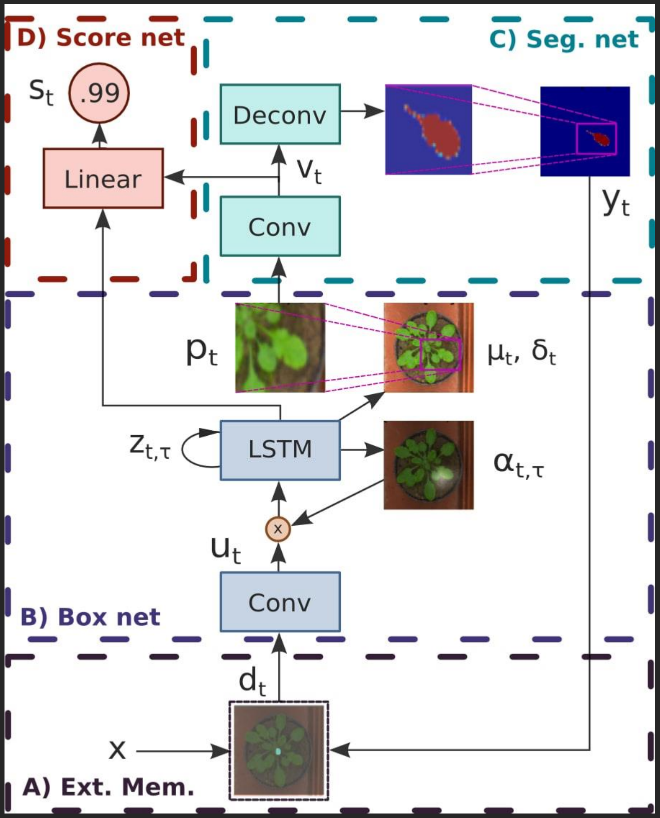
Sotirios A. Tsaftaris, PhD

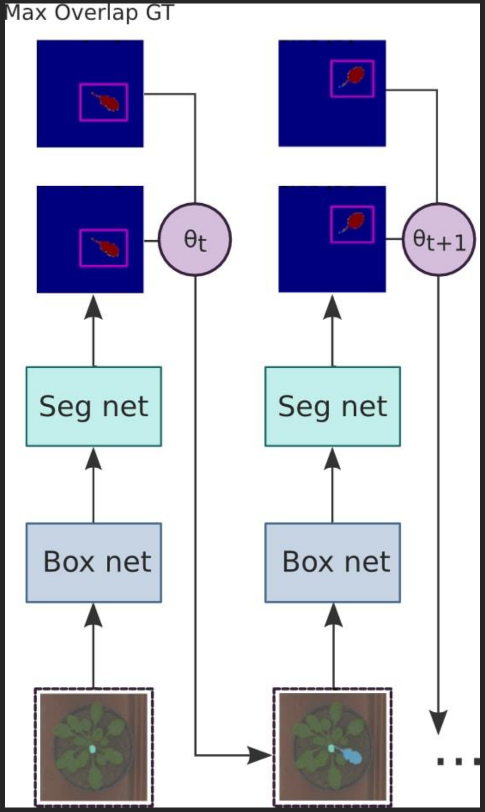
Email: S.Tsaftaris@ed.ac.uk

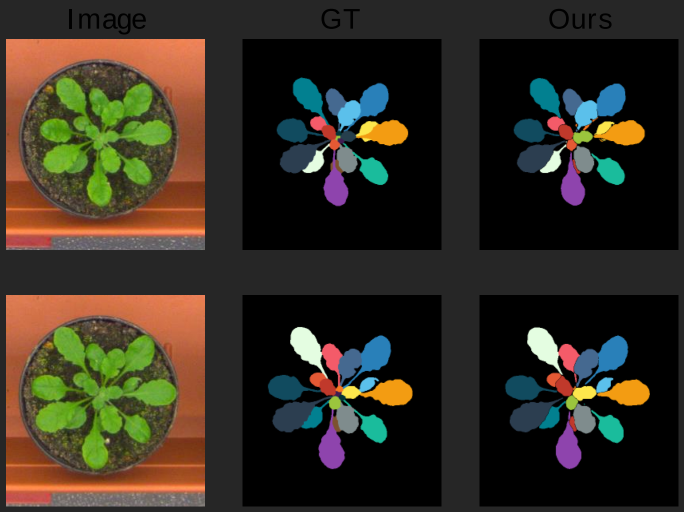
URL: http://tsaftaris.com

Leaf segmentation with recurrent neural nets

- Impressive segmentation/counting accuracy







	SBD ″	|DiC| #
RIS+ CRF [19]	66.6 (8.7)	1.1 (0.9)
MSU [20]	66.7 (7.6)	2.3 (1.6)
Nottingham [20]	68.3 (6.3)	3.8 (2.0)
Wageningen [26]	71.1 (6.2)	2.2 (1.6)
IPK [14]	74.4 (4.3)	2.6 (1.8)
PRIAn [6]	-	1.3(1.2)
Ours	84.9 (4.8)	0.8 (1.0)

Ren & Zemel End-to-End Instance Segmentation and Counting with Recurrent Attention arXiv